

OCCF 문제 해결을 위한 메타 정보 기반 가중치 결정 전략

A Weighting Scheme Using Meta Data for One-Class Collaborative Filtering Problem

저자 (Authors)	박주안, 황원석, 김상욱 Juan Parc, Won-Seok Hwang, Sang-Wook Kim
출처 (Source)	한국정보과학회 학술발표논문집 , 2014.6, 225-226(2 pages)
발행처 (Publisher)	한국정보과학회 The Korean Institute of Information Scientists and Engineers
URL	http://www.dbpia.co.kr/journal/articleDetail?nodeId=NODE02443978
APA Style	박주안, 황원석, 김상욱 (2014). OCCF 문제 해결을 위한 메타 정보 기반 가중치 결정 전략. 한국정보과학회 학술발표논문집, 225-226
이용정보 (Accessed)	한양대학교 166.***.182.218 2020/09/08 14:51 (KST)

저작권 안내

DBpia에서 제공되는 모든 저작물의 저작권은 원저작자에게 있으며, 누리미디어는 각 저작물의 내용을 보증하거나 책임을 지지 않습니다. 그리고 DBpia에서 제공되는 저작물은 DBpia와 구독계약을 체결한 기관소속 이용자 혹은 해당 저작물의 개별 구매자가 비영리적으로만 이용할 수 있습니다. 그러므로 이에 위반하여 DBpia에서 제공되는 저작물을 복제, 전송 등의 방법으로 무단 이용하는 경우 관련 법령에 따라 민, 형사상의 책임을 질 수 있습니다.

Copyright Information

Copyright of all literary works provided by DBpia belongs to the copyright holder(s) and Nurimedia does not guarantee contents of the literary work or assume responsibility for the same. In addition, the literary works provided by DBpia may only be used by the users affiliated to the institutions which executed a subscription agreement with DBpia or the individual purchasers of the literary work(s) for non-commercial purposes. Therefore, any person who illegally uses the literary works provided by DBpia by means of reproduction or transmission shall assume civil and criminal responsibility according to applicable laws and regulations.

OCCF 문제 해결을 위한 메타 정보 기반 가중치 결정 전략*

박주안⁰¹, 황원석², 김상욱²¹한양대학교 컴퓨터 소프트웨어학과²한양대학교 전자컴퓨터통신학과

{lmntdia, hws23, wook}@agape.hanyang.ac.kr

A Weighting Scheme Using Meta Data for One-Class Collaborative Filtering Problem

Juan Parc⁰¹, Won-Seok Hwang², Sang-Wook Kim²¹Dept. of Computer Software, Hanyang University²Dept. of Electronics and Computer Engineering, Hanyang University

요 약

One-Class Collaborative Filtering 문제를 해결하기 위한 방안이 제안되었으나, 누락 평점의 가중치는 단순한 전략에 의해 결정되었다. 본 논문은 누락 평점의 가중치를 보다 정교하게 결정하기 위하여 메타정보를 이용하는 전략을 제안한다. 실험을 통하여 메타정보를 이용하는 가중치 결정 전략이 기존의 전략보다 정확도를 더 향상시키며, 더 다양한 메타 정보를 이용할수록 정확도가 더 향상됨을 보인다.

1. 서 론

협업 필터링 (collaborative filtering)은 유저의 아이템에 대한 선호도를 예측하기 위하여 널리 이용되는 접근 방안이다 [1]. 이를 위해, 유저의 클릭, 북마크 등과 같은 행동을 이용하는 CF 기반 추천 방법들 [2]이 제안되었다.

유저의 행동을 이용하는 경우 유저가 아이템에 남긴 행동은 그가 해당 아이템을 선호함으로 간주할 수 있으나, 행동을 남기지 않은 아이템에 대한 유저의 선호도는 모호하다. 예를 들어 유저가 구매한 아이템은 그가 선호한다고 추측할 수 있다. 그러나 구매하지 않은 아이템은 (1)아이템을 선호하지 않아 구매하지 않았을 수 있으며 (negative example), 또는 (2)해당 아이템을 선호하지만 그 아이템을 모르기 때문에 구매하지 않았을 수 있다 (missing positive example). 이처럼 유저의 행동이 없는 누락 평점 (missing rating)에 대한 해석이 모호한 상황을 *one-class collaborative filtering (OCCF)* 문제라 한다 [3].

기존 연구 [3]는 누락 평점을 negative example로 간주하고, 누락 평점에 가중치를 부여함으로써 missing positive example일 가능성 또한 고려하였다. 이 때, 누락 평점의 가중치는 *유니폼*, *유저 기반*, *아이템 기반 전략*으로 계산하였다. 유니폼 전략은 모든 누락 평점의 가중치를 동일한 값으로, 유저 기반 및 아이템 기반 전략은 각각 유저별, 아이템 별로 누락 평점의 가중치를 동일한 값으로 부여한다. 그러나 동일한 유저 (아이템)에 대한

‘누락 평점들의 negative example일 가능성’은 다를 수 있으므로 정교한 가중치 결정 전략이 필요하다.

본 논문에서는 메타 정보를 추가적으로 이용하여 누락 평점의 가중치를 결정하는 새로운 전략을 제안한다. 아이템의 메타 정보는 유저의 구매 (클릭) 여부에 영향을 주는 요인 중 하나이다. 따라서 특정 유저가 구매 (클릭)한 아이템과 메타 정보가 유사한 아이템들은 구매 (클릭)할 가능성이 높다. 또한, 더 많은 종류의 메타 정보를 이용하면 누락 평점의 가중치 또한 더 정확하게 계산될 것이다. 본 논문에서는 실험을 통해 제안하는 가중치 결정 전략이 정확도를 향상시키며, 더 많은 종류의 메타 정보를 이용할수록 정확도가 더 향상됨을 보인다.

2. 제안하는 추천 방법

OCCF 문제를 해결하기 위하여 이용되는 추천 모델은 weighted low rank approximation (wLRA)에 기반한다 [3]. 이 모델은 평점 행렬과 함께 가중치 행렬을 이용하여 유저의 아이템에 대한 선호도를 예측한다. 평점 행렬의 인자는 유저가 아이템에 행동을 남기면 1이 부여되고, 그렇지 않은 경우는 누락 평점을 나타내며, 0을 부여함으로써 잠정적으로 negative example로 간주한다. 가중치 행렬에서 값이 1인 인자는 유저가 아이템에 행동을 남긴 경우이며, 누락 평점에 해당하는 인자는 가중치 결정 전략에 의해 값이 결정된다. 이 가중치 행렬에서 인자의 값이 클수록 추천 모델 학습 시 평점 행렬의 인자 값은 더 많이 반영된다.

wLRA는 평점 행렬을 목적 함수에 맞도록 분해함으로써 잠재적 요인에 대한 유저와 아이템의 특성을 추출한다. 특정 유저의 아이템에 대한 선호도는 이 유저와 아이템의 특성을 조합하여 결정된다. 목적 함수는 수식 1로 표현되며, 이를 구하기 위한 방법으로 weighted Alternating Least Squares가 이용된다.

* 본 연구는 미래창조과학부 및 정보통신산업진흥원의 대학 IT연구센터 지원사업(NIPA-2013-H0301-13-4009)과 문화체육관광부 및 한국저작권위원회의 2013년도 저작권기술 개발사업의 연구결과로 수행되었음. 또한 정보통신산업진흥원의 IT/SW 창의연구과정의 연구결과로 미래창조과학부와 마이크로소프트연구소에 의해 지원된 과제(NIPA-2013-H0503-13-1029)로 수행되었음.

$$\mathcal{L}(X, Y) = \sum_u \sum_i W_{u,i} \{(R_{u,i} - X_u \cdot Y_i^T)^2 + \lambda(\|X_u\|_F + \|Y_i\|_F)\} \quad (1)$$

수식 1에서 $R_{u,i}$ 와 $W_{u,i}$ 는 유저 u 와 아이템 i 에 대한 평점 행렬과 가중치 행렬의 인자를 각각 나타낸다. X_u 과 Y_i 는 각각 사용자와 아이템의 잠재적 요인에 대한 특성을 나타낸다. $\|\cdot\|_F$ 는 Frobenius 정규화를 나타내는 기호이며, λ 는 조정 파라미터를 의미한다.

제안하는 가중치 결정 전략에서는 유저가 행동을 남긴 아이템들과 유사한 메타 정보를 지닌 아이템들에 유저가 행동을 남길 가능성이 높다고 생각한다. 즉, 누락 평점은 그 평점에 해당하는 아이템이 기존 유저가 행동을 남긴 아이템들과 유사할수록 missing positive example에 가까우며, 유사하지 않을수록 negative example로 가깝게 간주된다. 누락 평점의 가중치는 negative example에 가까울수록 큰 값이 부여되며, 수식 2에 의하여 계산된다.

$$W_{u,i} = \begin{cases} 1 - \frac{\sum_{j \in C_u} \text{sim}(i, j)}{|C_u|} & \text{if } R_{u,i} \text{ is a missing rating,} \\ 1 & \text{otherwise} \end{cases} \quad (2)$$

수식 2에서 $W_{u,i}$ 는 유저 u 와 아이템 i 에 대한 가중치를 의미하며, C_u 는 유저 u 가 행동을 남긴 아이템 집합을 의미한다. 또한 $\text{sim}(i, j)$ 는 아이템 i 와 j 의 메타 정보에 대한 유사도이며, Jaccard coefficient로 계산한다. 그림 1은 제안하는 가중치 결정 과정에 대한 예이다.



그림 1. 메타 정보를 이용하는 가중치 결정 과정.

3. 성능 평가

OCCF 문제 해결을 위한 wLRA 모델에서의 가중치 결정 전략에 따른 정확도를 분석하기 위하여 본 실험에서는 Movielens 데이터를 이용하였다. 해당 데이터에는 943명의 유저와 1682개의 아이템(영화)이 포함되어있으며, 영화의 개봉 시기와 장르가 메타 정보로 존재한다. 또한, 유저가 아이템을 평가한 정보가 100,000개 있으며, 본 실험에서는 평가 여부를 유저의 행동으로 이용하였다.

정확도 평가를 위하여 평가 정보 중 80%를 트레이닝 세트로, 나머지 20%를 테스트 세트로 구분하였다. 트레이닝을 통해 평점 예측을 수행하고, 높은 선호도로 예측된 아이템 중 테스트 세트에 포함된 아이템의 비율을 precision과 recall을 통해 보였다. wLRA 수행을 위하여 Graphchi* 오픈 소스를 이용하였으며, 관련된 파라미터는 오픈 소스의 디폴트 값으로 고정하였다. 비교를 위해 OCCF 문제를 다룬 논문에서 제안한 유저 기반 전략의

정확도를 측정하였으며, 제안하는 전략에서는 메타 정보로 '장르'만을 이용한 경우와 '장르'와 '출시 연도'를 함께 이용한 경우의 정확도를 각각 측정하였다.

그림 2와 3은 wLRA에서 가중치 결정 전략들에 따른 precision과 recall을 보인 결과이다. 각 그림에서 x축은 예측된 선호도를 통해 추천한 아이템의 수를 의미한다. 기존의 유저 기반 전략을 적용한 경우보다 메타 정보를 이용한 전략들이 더 높은 precision과 recall을 보였으며, 가중치 결정에 더 많은 종류의 메타 정보를 이용할수록, 추천 정확도가 더 향상되었다. 따라서 본 실험에서 사용하지 않은 더 많은 종류의 메타 정보를 이용한다면 정확도를 보다 향상시킬 수 있을 것으로 기대한다.

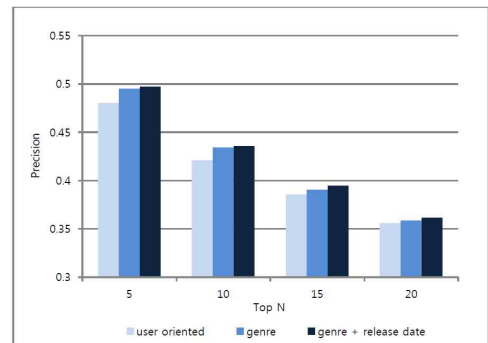


그림 2. 가중치 결정 전략에 따른 precision.

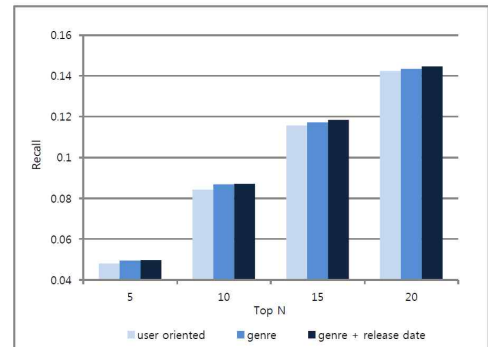


그림 3. 가중치 결정 전략에 따른 recall.

4. 결 론

본 논문은 메타정보를 이용하여 누락 평점의 가중치를 보다 정교하게 결정하는 전략을 제안하였다. 제안하는 전략에서는 유저가 이전에 선호한 아이템들과 유사한 메타 정보를 지닌 아이템들을 역시 선호할 것이라는 가정을 기반으로 아이템의 메타 정보의 유사도 계산을 통해 가중치를 결정한다. 실험을 통해 메타정보를 이용한 가중치 결정 전략이 기존의 전략보다 더 높은 정확성을 보임을 확인하였으며, 더 다양한 메타 정보를 이용할수록 최종 추천의 정확도가 향상됨을 보였다.

참고 문헌

- [1] D. Goldberg et al., "Using Collaborative Filtering to Weave an Information Tapestry," *Communications of the ACM*, Vol.35, No.12, pp.61-70, 1992.
- [2] D. Kelly and J. Teevan, "Implicit Feedback for Inferring User Preference: A Bibliography," *ACM SIGIR Forum*, Vol.37, No.2, pp.18-28, 2003.
- [3] R. Pan et al., "One-Class Collaborative Filtering," In *Proc. IEEE Int'l Conf. on Data Mining, ICDM*, pp.502-511, 2008.

* <https://github.com/GraphChi>